

A Posteriori Mathematics, Calculators and Malament-Hogarth Space-Times

April 13, 2011

Abstract

Are all mathematical facts *a priori*? I will argue that if two common but esoteric-sounding assumptions hold, some mathematical facts are only knowable *a posteriori*. The two assumptions are: that there are true but *a priori* unknowable Π_1^0 sentences, and that Malament-Hogarth space-times are metaphysically possible.

1 Introduction

By standard definitions, *a priori* knowledge is knowledge that does not depend on experience for its justification. And *a priori* truths are truths which one could in principle have *a priori* knowledge of. In contrast *a posteriori* truths are truths which can be known only in ways that depend on experience of justification. Are any mathematical truths *a posteriori* in this sense?

In this paper I will argue that if two common but esoteric-sounding assumptions hold, there are *a posteriori* mathematical truths. The two assumptions are: that some true Π_1^0 sentences are not provable from anything we are justified in believing *a priori*, and that Malament-Hogarth space-times are metaphysically possible. I will begin by arguing for the weak claim that particular bits of mathematical knowledge (like knowledge gleaned from looking at calculators) can depend on experience for their justification. Then I will propose a strategy for arguing from the two assumptions above to the stronger claim that certain kinds of physics experiments could give us justified belief in mathematical truths which we could never be justified in accepting *a priori*. Finally, in accordance with the strategy, I will go through a two step argument that takes us from these two premises to the conclusion that there are *a posteriori* mathematical facts.

2 A posteriori knowledge of mathematics

Everyone agrees that our knowledge of mathematical facts can ‘depend on experience’ in some sense. You can learn mathematical truths by hearing testimony or by looking at a calculator. Arguably knowledge by testimony has some special features¹, so let me focus on the case of forming new mathematical beliefs by looking at a calculator.

In this section I will argue that, as claimed by Kripke in Naming and Necessity, knowledge gleaned from a calculator can depend on experience for its justification, and hence count as genuinely *a posteriori* knowledge. When someone learns that, say, the square root of 37 is 6.08276253 by looking at a calculator, their newfound knowledge can depend for its justification on *a posteriori* beliefs about how the calculator was constructed, the laws of electricity etc. Thus their belief is not only caused by sensory experience but depends on these intermediate scientific beliefs for justification. As a result it too counts *a posteriori*.

To simplify things, let’s suppose that you designed and built the calculator. Thus, there’s no question of the the calculator transmitting testimony from the person who built it to you. You know by ordinary mathematical reasoning that a certain algorithm correctly calculates square roots. And you also know certain things about electronics empirically: how current would flow through certain wires, if certain buttons were pushed, and how this would lead to a certain display lighting up the screen. Combining these two pieces of knowledge, you built a physical machine such that the facts about what the calculator *would* display if certain buttons were pushed systematically co-vary with the facts about *what the square root function yields* when applied to certain numbers, and you know that this is the case. Thus, you have strong reason to believe that, if you push the buttons labeled (say), “sqrt”, “37”, and then “=”, the result will be an inscription of the first 9 digits of the square root of 37. Thus, when you *do* push these buttons and the number “6.08276253” shows up, you are justified in believing that the answer will be correct.

Now I claim that in this situation your *a posteriori* beliefs about electricity play a crucial role in *justifying* your belief about the square root. For consider a situation where your beliefs about the electronics of the calculator are wrong. For example, suppose that unknown to you the machine is wired up to always print out “6.08276253” when asked for the square root of something, but this just happens to be the right answer to the first question

¹Burge ‘Content Preservation’ Philosophical Review 1993

you enter. In such a situation, you would not count as knowing that the square root of 37 is 6.08276253. At best you would be in a Gettier case, believing a true claim about math for reasons that depend on justified, but false, beliefs about how the machine in front of you works. The fact that you could *in principle* derive the relevant mathematical claim without looking at the calculator does nothing to justify your token belief, or help it count as knowledge in this case. Thus (I claim), your justification for believing that the square root of 37 is 6.08276253 crucially runs through your scientific beliefs about electronics. The latter are justified by appeal to experience, so your knowledge of the square root of 37 is *a posteriori*.

I think this is already an interesting - and perhaps ultimately underappreciated - point. For example, one of Tyler Burge's main motivations for taking testimony to have a distinctive epistemic status (it's supposed to directly transfer justification from one mind to another) seems to be to avoid the conclusion that there can be *a posteriori* knowledge of mathematical facts by saying that what looks like *a posteriori* learning is really just the transfer of justification via testimony. But this conclusion is undermined by the above-mentioned reasoning about calculators. There is nothing like testimony involved in this case. The person who built the calculator (your earlier self) didn't have the knowledge that 37 is 6.08276253, so they certainly didn't transmit this knowledge to you, by way of the calculator. And yet, surely, you would count as having knowledge. So, we will still have to admit that there can be *a posteriori* knowledge - whatever we say about the status testimony.

But Kripke's claim that that particular token mathematical beliefs are knowable *a posteriori* falls far short of the claim that there are *a posteriori* mathematical facts. For a proposition is *a priori* iff it can be justified without appeal to experience. And in the case of the square root of 37, someone could nonetheless acquire justification for this belief which did not depend on experience by going through the whole algorithm for calculating the square root in their head.

In contrast, I will argue that our two assumptions (Malament-Hogarth space times are metaphysically possible, and there are *a priori* unknowable Π_1^0 sentences) yield the stronger conclusion that there are mathematical claims which could be known *a posteriori* but could not be known *a priori*. That is, not only is there *a posteriori* knowledge of mathematical facts, but there are *a posteriori* mathematical facts.

3 A posteriori mathematical facts? A strategy

Now let's start in on the more ambitious project. My strategy has two steps. The first step is to show that our two assumptions combine to yield the conclusion that it would be metaphysically possible to build a machine which behaves as follows: it turns on a light within 5 minutes if a certain *a priori* unknowable mathematical statement S is false, and leaves the light unilluminated if the statement was true.

The second step will be to argue that a creature in this situation could rationally come to believe that they *had* built such a machine, and hence justifiably form the belief that S when they saw the light remain dark after 5 minutes. If the latter claim is true, it follows that S is knowable *a posteriori* despite not being knowable *a priori*. Thus S is a genuinely *a posteriori* mathematical fact.

Note that this is not to say that *we* could learn S *a posteriori*. There is reason to think we do not live in a Malament-Hogarth space time. So it might well be that S is permanently unknowable to us because the physics of our world does not allow for the right kinds of experiments. Nonetheless, S is still an *a posteriori* truth, in the sense that S cannot be learned without appeal to experience but could *in principle* be learned by appeal to experience (i.e., there are some metaphysically possible worlds in which people do learn that S *a posteriori*).

4 Step 1: The metaphysically possible machine

Our first assumption was that there are some true, but *a priori* unknowable, Π_1^0 sentences. Π_1^0 sentences are sentences of the form $\forall x F(x)$ where the quantifier ranges over the numbers, and the property F is one that can be recursively checked i.e. you could program a computer to check whether any given number has the property or lacks it, and this computer would be certain to return one answer or the other in a finite amount of time. So, Π_1^0 sentences say 'every number has the property F' where F is some property which it is checkable whether any particular number has or lacks.

Why think that there are true but *a priori* unknowable Π_1^0 sentences? It's by no means uncontroversial that there are such sentences. However it is commonly assumed that there are, for the following reason. *If* you think that the whole system of justified *a priori* reasoning can be recursively axiomatized, then the first Incompleteness Theorem yields the conclusion that there are true but *a priori* unknowable Π_1^0 sentences. The first Incom-

pleteness Theorem says that that no recursively enumerable system which captures certain basic facts about arithmetic can prove all true Π_1^0 sentences. In particular, it shows that that every such system S won't be able to prove the Π_1^0 sentence 'every number fails to code for a proof of G', where G is a specially-cooked up claim about arithmetic called the Gödel sentence for S. Thus *if* there is a definite body of justified *a priori* reasoning which could be formalized in such a way that a computer could in principle be programmed to check whether any given inference was acceptable, *then* there will be a true Π_1^0 sentence *a priori* reasoning cannot lead one to. Thus, there will be a true Π_1^0 sentence, which cannot be known *a priori*.

Now I am going to argue that this claim combines with our second hypothesis, to yield the conclusion that it would be metaphysically possible to build a machine that stays dark at the end of a specific 5 minute window if and only if a certain *a priori* unknowable mathematical statement is true. The first hypothesis says there's at least one a true Π_1^0 sentence that can't be arrived at via justified *a priori* reasoning, and hence is unknowable *a priori*. So let's call this sentence S. S says that every number has a certain property P. And we can check with pen and paper that, say, 3 has the property P, or with an ordinary computer that the first billion numbers have property P. But S makes a claim about infinitely many different numbers, so no finite amount of checking will suffice. At any finite time all we know is that we haven't come up with a counter-example yet. Normally we would try to come up with a mathematical argument that every number has to have property P, and which uses mathematical induction to secure the desired result for all the numbers. But by hypothesis there is no justified *a priori* argument which leads to the conclusion that S.

What we'd like to do would be to, somehow, check infinitely many cases (or build a machine that could do that). This is where the second hypothesis - that Malement-Hogarth space times are metaphysically possible - comes in. It has been shown that certain solutions to the equations for general relativity would allow a person and an ordinary computer to take different paths through space-time in such a way that information from infinitely many computations in the computer reach the person within what is (for them) a finite amount of time.²

Hogarth draws attention to these space-times because a computer user who took such a path relative to the computer would be able to "compute" things that a Turing machine cannot. In particular, a computer user in

²See 'Deciding Arithmetic Using SAD Computers', Mark Hogarth, British Journal of Philosophy of Science 2004

the situation above would be able to determine whether a regular turing machine halted. They would just set the computer at the other end to go through each stage of the computation, and then send a signal at stage n if and only if the computer halted at stage n . Thus, in particular, they could evaluate any Π_1^0 sentence (which says that every number has the recursively checkable property F), by programing the regular computer at the other end to go through the numbers in order, checking each and sending a signal only if it found a counter example.

Hogarth introduces these kinds of machines because he thinks that computer + user systems of the kind above would count as doing computations, so that whether or not this kind of set up is physically possible will make a difference to what is and isn't computable. Thus, he argues, it's a contingent matter of physics whether the Church-Turing thesis (that something is computable iff some Turing machine computes it) is false. But for our purposes it doesn't matter whether the operations in question would count as computations, or whether they are physically possible. Our second hypothesis is simply that the set-up Hogarth describes is metaphysically possible.

Hoagarth's setup involves two things: An MH space-time and something that behaves like a regular computer minus the fact that a regular computer would eventually break down. Suppose that both of these are metaphysically possible and can be combined (I take it that the metaphysical possibility of the computer that goes through "infinitely many stages" of computation without breaking is uncontroversial). Then it would be possible to set such a computer to check through each of the infinitely many possible counter-examples to our unknowable Π_1^0 sentence S , and send a signal to us within a certain 5 minute window if and only if it found a counter example. If the computer were programed to check instances of some false Π_1^0 sentence, there would be some stage at which the computer found a counter example and sent a signal. Let's say the computer sends the signal to some kind of panel, which lights up if it ever gets a signal but otherwise stays unlit. Thus if the panel stays unlit whatever Π_1^0 sentence which the computer was programed to check is true.

This completes the first stage in my argument: it would be metaphysically possible to build a machine which would turn a light on if an *a priori* unknowable mathematical statement S is true, and not if this sentence is false.

5 Step 2: Knowledge about the machine

But this doesn't yet suffice to show that S is knowable *a priori*. If a subject knew that he was dealing with a machine of the kind just described, and he saw that the light did not turn on, he would be justified in believing that S sentence was true. But could someone ever know that they were in the kind of set up Hogarth describes?

Establishing this point (that a rational subject could *learn* that the physical system would behave a certain way if S is true) is quite crucial to the argument, and the difficulty of establishing this point explains why I've chosen such a complicated example. For if we just wanted to show that it would be metaphysically possible for there to be a system whose state systematically reflected *a priori* unknowable mathematical facts, no elaborate description is needed to establish this. It would clearly be metaphysically possible for there to be an oracle which simply spit out the right answer to any mathematical query. There could just be some black box, and a fundamental law of physics which said that that's how the box behaved.

However, with these examples it's not at all clear what *evidence* could rationally convince you that this really was an oracle. You could check some finite number of predictions from the box, but simply knowing that some physical system gets all the cases you have checked right doesn't seem sufficient to justify the conclusion that it gets everything right, in the absence of any plausible mechanism for how these correct predictions might be generated. And, arguably, you would never have any plausible mechanism for how the oracle works since in this scenario it's just a basic physical law that the oracle's answers co-vary with the mathematical facts in this way.

In fact, even more simply, the literal requirement above simply said that the system had to go into a certain state iff a single target sentence S was true. But, actually, *every* object has that property. Since (by stipulation) S is true, ripe bananas have the property of being yellow if and only if S is true, rocks have the property of being attracted to the earth iff S is true etc. But considering just these cases, there is no reason to think someone could ever learn that the above biconditional holds. (Remember that S is supposed to be unprovable so they can't directly learn that S is true by *a priori* reasoning, and then infer these sentences of the form, 'if X then S '.)

Thus even if it's obvious that it would be metaphysically possible to have a system which goes into a certain state if and only if a certain Π_1^0 sentence was true, it's not clear that you could be in a position to *know* this. This is where the exotic machinery of experiments in a Malament-Hogarth space-time come in. For, I will argue, someone in this situation very plausibly

could get sufficient empirical evidence to justify belief that they were in a Malament-Hogarth space-time, and stood in the relevant relationship to a computer that would check all cases of S.

Admittedly there is no course of experience which would logically entail that the physical world around one is set up in the relevant way. But this is not the standard which we generally require for knowledge of scientific facts like the structure of space time, or the future behavior of a certain computer. I take the existence of actual evidence-heavy debates in physics about whether the best laws of physics as currently known are compatible with our being in a Malament-Hogarth space time to be strong evidence that some course of experience could rationally lead one to believe that one was in the relevant situation. And, more generally, I take the history of physics to provide examples of the kind of evidence which might rationally convince someone that space-time had a certain structure.

However, even granting that the a space-time with the relevant structure is possible, there might be worries about the long-lived computer needed to create Hogarth's set up. Certainly a computer which goes through each stage of the computation and never breaks seems metaphysically possible, but would it ever be rational to believe that one was dealing with such a machine?

Here, I want to suggest that it could look like a simple consequence of basic physical laws (the kind one would rationally expect to be preserved forever) that the machine in question behaved like this. We might have to imagine evidence that, contrary to what's actually the case, the universe allowed something like perpetual motion and that energy could be created. But it does seem conceivable that one's best theory of the behavior of some kind of fundamental particle would be that, e.g. it emits a pulse of light every second, or it always behaves like an OR gate with respect to other particles associated with it in a certain way, and that one could use such particles to build the relevant kind of enduring computer.

In light of these considerations, it seems like a) someone could be in the situation that Hogarth describes and b) they could come to know that they were in this situation. Furthermore, if they were in this situation and they saw that the relevant light didn't come on, they could rationally infer from this experience, and the fact that they were in the relevant Hogarth set up, that S was true. Thus, it would be possible for someone to learn that S *a posteriori*. So, if our two hypotheses are true, there can be *a priori* unknowable facts which nonetheless could be learned *a posteriori* were contingent facts about nature to prove suitably cooperative.

6 Conclusion

In this paper I have argued that if a) there are *a priori* unknowable Π_1^0 sentences and b) Hogarth's set up is metaphysically possible, then there will be *a posteriori* mathematical facts. That is: there will be mathematical facts which could (in principle) be known, but only in ways that would depend on experience for their justification. If this is correct we are left with three options: reject a), reject b), or accept the conclusion that some mathematical facts are *a posteriori*. It's beyond the scope of this paper to say which of these options we should ultimately take, but I will end by briefly suggesting one reason for rejecting a).

It may seem like the only alternative to the conclusion that there are unknowable Π_1^0 facts is to posit some mystical sense in which our minds are not mechanistic, so that our mathematical reasoning leads us to a set of conclusions which is not recursively enumerable. But in fact, there's a much simpler reason for rejecting a).

All we need to do, is 'sharply distinguish logic from psychology', by considering creatures who are inclined to accept quite different propositions from the ones that we currently accept. Some mathematical statements, like $2+2=4$ or the least number principle, strike us as obvious. We don't ask for proofs of such propositions, but rather use them to construct proofs of other statements. Now couldn't there be creatures with a different psychology from us, who found other mathematical truths equally obvious? And what kind relationship do we bear to the (recursively enumerable) mathematical truths that feel obvious to us, which these creatures would not bear to the (different but also recursively enumerable), mathematical truths that feel obvious to them? It's at least tempting to think that such a creature would count as having *a priori* knowledge of whatever conclusions it derives from the true mathematical principles which seem obvious to it, just as much as we do. And if this is right, then all mathematical facts (which we can form beliefs about) are knowable *a priori*.